基于分布式图计算的学术论文推荐算法 *

潘 峰，怀丽波†，崔荣一

(延边大学 工学院 计算机科学与技术学科 智能信息处理研究室, 吉林 延吉 133002)

摘 要：针对海量论文数据导致的应用效率低下问题，提出一个基于层次混合模型的推荐算法 WSVD++。该模型根据学术论文良好的结构特征，构建一个加权的论文二部图模型。首先对论文进行特征提取，按不同特征的权重构建论文的复合关系图；其次对关系图采用一种改进的 PPR 算法，计算每篇论文的重要程度，依此来对用户—论文关系进行加权；然后在构建好的加权二部图模型上混合 SVD++图算法进行推荐。实验结果表明，改善了推荐算法学术论文的推荐效果，并且基于分布式图计算框架 GraphX，扩展性好，适合大数据处理。

关键词：混合模型推荐；协同过滤；SVD++；分布式图计算；GraphX

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2018.01.0003

Academic paper recommendation based on distributed graph

Pan Feng, Huai Libo†, Cui Rongyi

(*Intelligent Information Processing Lab*, *Dept. of Computer Science & Technology*, *Yanbian University*, *Yanji Jilin* 133002, *China*)

Abstract: Aiming at the low efficiency caused by massive academic paper data, this paper proposed a recommendation algorithm method based on the hierarchical mixed model named WSVD++. According to the structural features of academic papers, the model constructs a weighted bipartite graph model. Firstly, this method extracted the features of each paper and constructs the composite relation graph according to the ratio of different features. Secondly, it uses an improved PPR algorithm on the graph to calculate the importance weight of each paper, and then weighs the relation between the user and the paper. Finally, it recommend on the weighted bipartite graph by using SVD++ graph algorithm. The result shows that the proposed algorithm improves the recommended accuracy. The whole process implemented in distributed graph calculation system, that means the method has good expansibility and is suitable for big data processing.

Key words: hybrid model; collaborative filtering; SVD++; distributed graph computation; GraphX

0 引言

学术论文作为学术思想传播的载体，在信息爆炸的时代中同样存在数据指数增长的问题。据 DBLP[1]统计，自 2010 年以来每年平均有超过 32.7 万篇论文发表，平均每年环比增长 5.56%，每年在期刊和会议发表的论文数量相当庞大。近年来检索系统性能的提升，减轻了科研工作者查找学术论文的工作。但是面对会议和期刊每年产生的海量数据，论文检索系统已经不能满足科研工作者的需求。对此，学术论文推荐系统可以解决这个问题[2]。

在推荐算法领域，受到 Netflix 和 KDD Cup 等竞赛的推动，协同过滤等推荐算法得到了人们的普遍关注并取得了很大的进展。而学术推荐应用领域，仅在 CiteULike，ResearchGate 等文献管理网站中有所涉及。论文推荐属于一类协同过滤问题（OCCF）[3]，对于传统推荐技术很难直接应用。文献[4]通过引入权威作者等论文信息来深化用户的兴趣模型从而推荐高质量论文。文献[5]利用主题关系提取用户和论文的主题模型从而缓解冷启动问题。论文的数据结构不同于一般推荐的项目，通常为非结构化或半结构化的数据，因此采用图模型的结构可以对丰富的信息进行建模[6]，从而更好的表达论文关系。文献[7]通过共引分析提出基于 DBSCAN 的密度聚类算法来找到相近论文。文献[8]通过对比不同算法应用到图模型中的表现，引入论文的引文和内容信息提出了一种异构图推荐方法。面对海量增长的数据，大规模的图计算的任务处理是至关重要的[9]，对于海量增长论文推荐同样亟待解决的问题。

在本文中，对于论文的结构特征和 OCCF 导致的推荐算法难以直接应用的问题，提出一种基于分布式图计算的论文推荐方法。

收稿日期：2018-01-03；**修回日期**：2018-02-13　　**基金项目**：国家语委“十二五”科研规划 2015 年度科研项目（YB125-178）

作者简介：潘峰（1993-），吉林汪清人，男，硕士研究生，主要研究方向为数据挖掘、推荐系统；怀丽波（1973-），黑龙江五常人，女（通信作者），副教授，主要研究方向为优化理论与方法、数据挖掘（huailibo@ybu.edu.cn）；崔荣一（1962-），吉林延吉人，男，教授，主要研究方向为模式识别、智能计算.

1　相关算法介绍

1.1　基于内容的协同过滤

论文推荐常采用基于内容的协同过滤算法[10]。给定论文集合 I，通过用户 u 已经操作的论文 $i \in I$ 来确定带有共同特点的候选论文 $\hat{r}_{ui}$。首先计算论文 i 中每个词项 t 的加权词频逆文档频率（$wf\text{-}idf_t$），如式（1）所示。

$$wf\text{-}idf_t = \left(1+\log tf_t\right)\cdot \log \frac{|I|}{N(t)} \tag{1}$$

其中：tf_t 表示 t 在论文中的频率，$N(t)$ 表示 t 在 I 中出现的次数。然后通过论文之间的相似度关系来计算 $\hat{r}_{ui}$。如式(2)所示。

$$\hat{r}_{ui} = \sum_{j \in R(u)\cap S(i,K)} w_{ij} r_{uj} \tag{2}$$

其中：$R(u)$ 表示包含用户 u 评分的所有物品集合，$S(i,K)$ 表示和论文 i 最接近的 K 个论文的集合，w_{ij} 表示论文 i 和论文 j 的余弦相似度，如下所示：

$$w_{ij} = \frac{|i \cdot j|}{\sqrt{|i|\cdot|j|}}$$

1.2　基于模型的协同过滤

基于模型的协同过滤是构建隐语义模型（LFM）[11]，将用户和论文映射到一个维数为 f 的低维联合语义空间中。在 LFM 中，用户特征向量和论文特征向量的内积来反映用户的兴趣程度，即 $\hat{r}_{ui} = q_i^T p_u$。其中 $p_u \in R^f$ 和 $q_i \in R^f$ 分别表示用户和论文的隐因子向量。

通过基准预测和隐式反馈信息可以增加预测准确度，即在 LFM 模型中加入偏置项 $b_{ui} = \mu + b_u + b_i$ 和带隐式反馈的用户因子向量 $p_u + |R(u)|^{-\frac{1}{2}} \sum_{i \in R(u)} y_i$。这种方法称为 SVD++模型[12]，如式（3）所示。

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T \left(p_u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j \right) \tag{3}$$

其中：μ 为总体平均分，b_u 和 b_i 分别表示用户 u 和论文 i 与 μ 的偏差。$y_i \in R^f$ 表示论文 i 的偏置向量，表示论文 i 与其他论文的误差。

1.3　基于图模型的推荐

图模型以几何方式展现数据关系，不仅可以通过信息传递计算直接连接的顶点关系，也可以计算非直接连接的顶点间的影响[13]。对于直接连接的顶点，边的权重越大，顶点的关系越密切。基于图模型的推荐是通过数据建立图模型，然后在图上驱动推荐算法产生候选推荐。

1.3.1 论文关系图构建

论文图模型推荐考虑两种直接关系来计算间接关系[14]。一种是用户和论文的直接关系，将用户论文关系数据表示为一个二部图 $G(V,E)$，其中 $V = V_u \cup V_i$，由用户顶点集合 V_u 和论文顶点集合组成 V_i。$E = (V_u, V_i, R)$，表示用户 u 对论文 i 的评分 r，每条边 e 都对应用户行为数据中的三元组 (u,i,r)。另一种是用户间或论文间的直接关系。如论文间的引用关系是一种类似于网页链接关系的有价值信息。若论文存在引用关系，则它们之间的紧密程度会强于其他论文。对于追本溯源的用户，引用关系可以体现特定领域在不同阶段的研究状况。

1.3.2 基于图模型的 PPR 算法

Personal PageRank（PPR）算法是 PageRank 算法应用到推荐的一种改进算法，以目标用户为出发点，在论文顶点和用户顶点之间随机游走[15]。游走过程会按概率 d 跳到下一步，并在多次游走后收敛。最终，候选推荐的论文通过权重扩散来计算：

$$PR(i) = d \cdot r + (1-d)\cdot \sum_{i' \in C(i)} \frac{PR(i')}{deg(i')} \tag{4}$$

其中：$C(i)$ 表示与顶点 i 连接的顶点集合，$deg(i')$ 表示顶点 i' 的出度（degree）。r 表示偏好向量取 0 或 1，当顶点 i 为用户顶点 v_u 时 r 为 1。

1.4　分布式图计算

随着图结构数据的规模和分布式图计算的重要性不断增加，出现了一系列图形分布式系统，如 Pregel、GraphLab 和 GraphX 等。分布式图计算是由一个稀疏图 $G(V,E)$ 和一个在每个顶点 $v \in V$ 上并行执行的顶点程序 Q 组成，每个顶点 v 同时被实例化为 Q (V)并且可以通过消息（Pregel）或共享状态（GraphLab，GraphX）与相邻的顶点程序交互[16]。分布式图计算系统抽象地约束了顶点程序与图形结构的交互，从而实现了数据布局和通信的优化。顶点程序 $Q(v)$ 可以读取和修改顶点 v 的属性以及相邻边 e 上的属性，甚至在某些情况下可以修改邻接顶点上的属性。

2　基于图模型的论文推荐算法

本文提出一种基于分布式图模型的学术论文推荐算法。基本思想如下：a）首先构建论文两层关系图。对论文内容进行特征提取，包括题目、摘要、引用关系，分别构建三个论文属性关系图；根据特征按不同权值比例合并图，并在合并的论文关系图中计算论文的重要程度；b）构建用户-论文两层图。用户层和论文层之间边的初始权值为用户对论文的操作记录，但由于用户的行为数据只有操作和未操作两种情况，很难表达用户对论文的偏好，因此利用论文内容属性信息，根据论文的重要程度对用户和论文的关系进行加强；c）得到融合内容和操作信息的二部图模型，利用图模型进行论文推荐。以上图模型的构建和计算均采用分布式的图操作。

2.1　图操作定义

图操作包括图的构建（makeG）、图顶点和边的属性更改（mapVertices 和 mapEdges）、图和图的合并（mergeG），外部信息融合（JoinVertices）以及图上消息传递（aggregateMsgs）。根据 GraphX 框架，这里针对本算法给出几种重要的操作函数定义。

定义 1　makeG。输入三元组数据建立不可变的图的索引结构 Graph[V,E]。其中 V 是 Vertex [Vid，Vattr]的缩写，由顶点

索引 id 和顶点属性构成；E 是 Edge [Vid, Vid, Eattr]的缩写，由源顶点 id、目标顶点 id 以及边属性构成。Graph 结构可以表示成 EdgeTriplet[V,E]的三元组形式，即((srcId, srcAttr), (dstId, dstAttr), Eattr)。

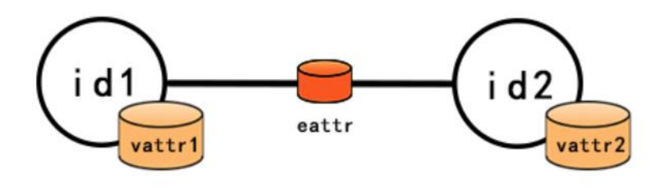


图 1　图模型的索引结构

定义 2　mapVertices 和 mapEdges。两个函数分别可以对应的修改 Graph 中的 Vattr 和 Eattr，并返回一个修改后的 Graph。

定义 3　mergeG 和外部信息融合 JoinVertices。mergeG 函数要求输入两个 Graph 和权值，将两个 Graph 中对应的 Eattr 按权值合并，顶点间无连接的边 Eattr = ∅，最后返回合并的 Graph。JoinVertices 可以使当前图接入外部 Vertex 信息，首先按照 Vid 进行连接，然后将两个被连接顶点的 Vattr 进行 mapVertices 操作，最后将连接后的 V、未连接 V 以及 E 构成的图作为返回。

定义 4　aggregateMsgs 函数。该函数接受 sendMsg，mergeMsg 两个函数作为参数。其中 sendMsg 以边为输入，对于图中每条边，可以选择向源顶点（toSrc）或目标顶点（toDst）发送一个“消息”(Message)。mergeMsg 函数将每个顶点接收到的“消息”进行聚集，并将结果 mapVertices 到相应的顶点 Vattr 上，最后返回顶点集合 V。

2.2　论文关系图

从原始数据中分别提取题目、摘要和引用关系三个论文的内容特征，并对题目和摘要进行 tf-idf 加权处理。然后分别计算三个特征下论文之间的余弦相似度，并用 makeG 构建三个论文关系图 abGraph、tiGrap 和 ciGraph。图中顶点为论文，边的权值用余弦相似度表示。之后用 mergeG 按权值合并构建加权的论文关系图 tacGraph。

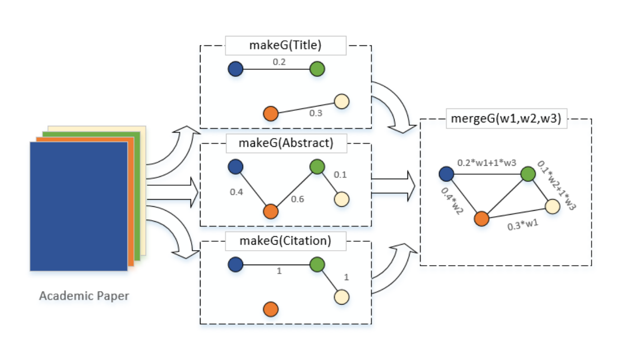


图 2　构建论文关系图

2.3　加权的用户-论文关系图

由于论文层中论文以相似度来表示关系程度，本文对 PPR 算法进行改进，加入论文相似度来计算论文重要程度。根据式（4）给定概率 d 作为抑制因子，对于顶点 i 的重要程度计算公式为

$$W(i)=d\cdot r+(1-d)\cdot\sum_{i'\in C(i)} w_{ii'} \tag{5}$$

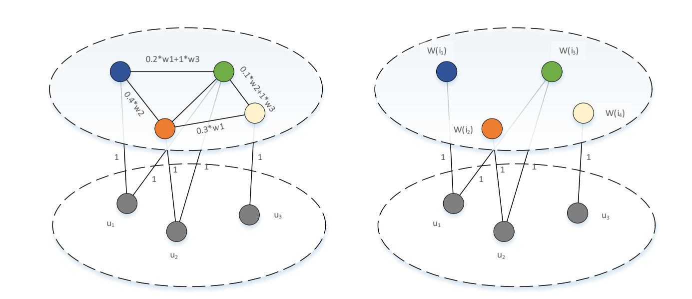
$$w_{ii'}=\frac{e_{ii'}}{|deg(i')|}\cdot W(i') \tag{6}$$

其中，$e_{ii'}$ 表示顶点 i 与 i' 关系的权值。

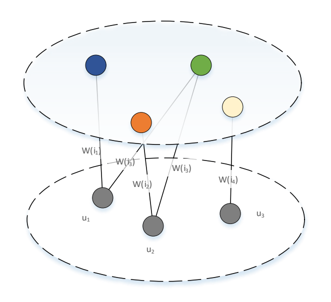
根据式（5），在论文关系图中可以找到比较重要的论文，即 W(i)值较高的论文。该图算法如下：

```
Algorithm 1: 论文图重要程度计算
Input: tacGraph, IterMax, d
1 W(i) = d
2 While i < IterMax
  VerticesWithW ← tacGraph aggregateMsgs (
    sendMsg: toDst( w_ii' )
    mergeMsg: MsgSum = Σ_{i'∈C(i)} w_ii' )//according Equation 6
  newtacGraph ← tacGraph JoinVertices(VerticesWithW)
  newtacGraph mapVertices (W(dst)) //according Equation 5
Output: newtacGraph
```

利用用户对论文的行为数据转换为用户-论文关系图 uiGraph，与 tac。并根据 Algorithm1 得到新的论文关系图 newtacGraph。然后用 newtacGraph 图中论文顶点的重要程度 W(i)替换 uiGraph 图中论文顶点的 Vattr。最后对 uiGraph 图中的边 Eattr 进行加强处理，即进行 mapEdges（$\mathrm{Eattr}\leftarrow\mathrm{W}(i)$）操作，得到新的二部图 WGraph。



(a)融合论文特征的用户-论文两层图　(b)确定论文节点权值的二部图



(c)将论文节点权值传递给边后的二部图

图 3　图模型的构建过程

2.4　基于分布式图模型的 SVD++算法

式（3）中的参数是通过采用随机梯度下降法（SGD）进行优化，损失函数用最小化相关联的正则化平方函数（SSE），如

式（6）所示。其中，$e_{ui}=r_{ui}-\hat{r}_{ui}$，常量 λ_1，λ_2 控制正则化程度。

$$\mathrm{SSE}=\frac{1}{2}\sum_{u,i}e_{ui}^2-\frac{1}{2}\lambda_1\left(b_u^{\ 2}+b_i^{\ 2}\right)+\frac{1}{2}\lambda_2\left(\left|p_u\right|^2+\left|q_i\right|^2+\sum_{j\in R(u)}\left|y_j\right|^2\right) \tag{7}$$

迭代过程每步更新如下。

$$\begin{cases} updateBu=\gamma\cdot\left(e_{ui}-\lambda_1\cdot b_u\right) \\ updateBi=\gamma\cdot\left(e_{ui}-\lambda_1\cdot b_i\right) \\ updateQ=\gamma\cdot\left(e_{ui}\left(p_u+\left|R(u)\right|^{-\frac{1}{2}}\sum_{j\in R(u)}y_j\right)-\lambda_2\cdot q_i\right) \\ updateP=\gamma\left(e_{ui}\cdot q_i-\lambda_2\cdot p_u\right) \\ updateY=\gamma\cdot\left(e_{ui}\cdot\left|R(u)\right|^{-\frac{1}{2}}\cdot q_i-\lambda_2\cdot y_j\right),\forall j\in R(u) \end{cases} \tag{8}$$

其中，γ 为步长。本文设置参数值为 $\gamma=0.007$，$\lambda_1=0.005$，$\lambda_2=0.015$ [12]。并且迭代过程设置衰减因子 α [17]，使每一次迭代的步长减少，在 30 次后收敛。

```
Algorithm 2：SVD++图法
Input: WGraph, f , IterMax
1  γ=0.007 , λ1=0.005 , λ2=0.015 //Initialize
2 svdGraph mapVertices( vattr={k1,k2,b=0,y=0|k1,k2∈R^f} )
//Initialize Vattr
3  Calculate  mean  //the average of Vattr
4  While i  <  IterMax
     Sumy  ←  WGraph aggregateMsgs (
       sendMsg：toSrc (Msg =  dst.k2 ),
       mergeMsg ：MsgSum  =  ∑Msg )
     svdGraph JoinVertices(Sumy)
     svdGraph mapVertices( {k1,k2=MsgSum,b,y} )
//Initialize  yi
     update ← svdGraph aggregateMsgs(
     Calculate  r̂ui   updateP ,  updateQ ,  updateY ,
 updateBu , updateBi //according Equation 3, 8
     sendMsg:
       toSrc
( {Msg1=updateP,Msg2=updateY,Msg3=updateBu} )
       toDst
( {Msg1=updateQ,Msg2=updateY,Msg3=updateBi} )
     mergeMsg:
( {Msg1sum=ΣMsg1,Msg2sum=ΣMsg2,Msg3sum=ΣMsg3)} )
       svdGraph JoinVertices( update )
     svdGraph
mapVertices( {k1+Msg1sum,k2+Msg2sum,b+Msg3sum,y} )
5  γ=α·γ
Output: svdGraph
```

3　实验结果与分析

3.1　实验数据与环境

本文使用 CiteULike 网站开源的数据集，原始数据包括 5551 名用户、16980 篇文章和 204986 条行为数据。本文选取其中稠密部分，并在数据集中抹掉 20%的用户的一半操作记录，以此作为训练集，抹掉的数据为测试集，如表 1 所示。

表 1　训练集和测试集

	用户数	论文数	记录数
训练集	848	1921	36215
测试集	173	1921	2596

实验环境选择小型 Spark 集群，该集群包括 4 个节点。每个节点 CPU 型号为 E3-1230v5，内存为 6GB。

3.2　评价标准

由于论文推荐属于 OCCF，本文选取准确率（Precision）、召回率（Recall）、F 值和覆盖率（Coverage）对混合模型进行评估。在测试集中，计算模型推荐的候选论文集合与用户实际操作的论文集合的交集作为正确的候选论文。

推荐列表长度为 M 的准确率和召回率定义为：

$$\mathrm{Precision@M}=\frac{\left|preset\cap testset\right|}{\left|testset\right|}$$

$$\mathrm{Recall@M}=\frac{\left|preset\cap testset\right|}{\left|preset\right|}$$

F 值又称为 F-score，是 IR（信息检索）中常用的一个评价标准，计算如下，通常取 $\alpha=1$。

$$F=\frac{\left(1+\alpha^2\right)Precision\times Recall}{\alpha^2\times Precision+Recall}$$

覆盖率测量的是推荐系统推荐给所有用户的物品数占总物品数的比例，如公式所示。

$$\mathrm{Coverage}=\frac{\left|\mathrm{U}_{u\in N(U)}\mathrm{R}(u)\right|}{\left|\mathrm{I}\right|}$$

3.3　实验结果分析

对于摘要、题目和参考文献的权重(w_1、w_2、w_3)，调节不同的分配会影响推荐的结果，如图 4 所示。其中，将非分配权重情况下（既 1:1:1）作为对比参考。

从图中可以看出，论文数据的摘要和引用关系特征对推荐结果的影响较为明显，都高于非分配权重情况。因此，根据调节的曲线，本文实验选择摘要、题目和引用关系的权值为 0.3、0.2 和 0.5。

对于式（5），使用不同的抑制因子对于结果会有不同的影响。文献[13]给出一般的 d 取 0.85。为了实验找出较合理的抑制因子，设置固定隐因子个数位 5，测试不同抑制因子下产生推荐结果，计算召回率和 F 值。

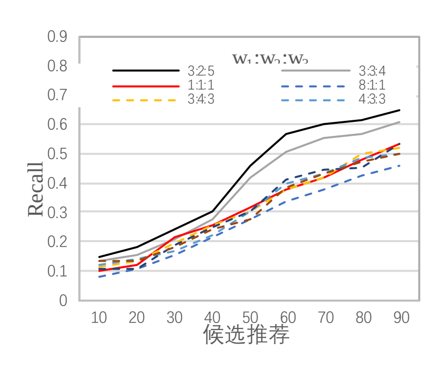


图 4　题目、摘要和参考文献不同权重下的召回率

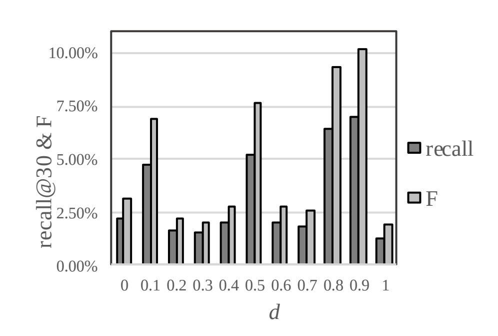


图 5　不同抑制因子 d 下混合推荐的召回率

由图可知，d 在选取 0.1、0.5、0.8、0.9 时召回率和 F1 值明显高于其他情况，这是由于在推荐领域用户个体主观兴趣不同所致。对于倾向经典的有价值的论文的用户，在论文关系中约趋于中心的论文越有价值，即该论文关联许多其他论文，因此选取区间[0.8,0.9]更合适；对于倾向冷门的新颖观点的论文，由于论文关联程度低，因此需要设置更高的跳转概率，即选取 0.1 附近更合适；而对于不确定的用户可以选取 0.5。本文其他实验抑制因子选取 0.9。

为了评估本文提出混合模型（WSVD++），对比实验选取基于内容（CB-CF）、基于引文（cite-PR）、基于 PPR 和基于 SVD++的图模型算法。其中，SVD++隐因子维度为 10，WSVD++隐因子维度分别选取 10 和 20，迭代过程设置衰减因子 $\alpha=0.9$，使迭代在 30 次后收敛。结果如表 2 所示。

表 2　实验结果

Approach	Recall@30	F	Coverage
CB-CF	18.73%	12.14%	49.55%
cite-PR	12.38%	6.86%	28.68%
PPR	7.45%	3.46%	22.44%
SVD++@10	11.02%	5.10%	25.40%
WSVD++@10	22.48%	17.03%	62.10%
WSVD++@20	23.16%	21.37%	63.87%

从表中可以看出，本文提出的混合模型推荐结果的召回率为 23.16%，F 为 21.37%，优于基于内容的协同过滤算法(CB-CF)、基于引文的图算法（cite-PR）、其他基于图模型的推荐算法(PPR 和 SVD++)。这说明用户操作记录结合了论文重要程度的权值，能够突出用户对重要论文的操作行为，从而建立的隐因子模型更贴近用户的个性化兴趣。并且，选取更长的隐因子也会对推荐结果有一定的提升。从覆盖率的角度考虑，混合模型结合了论文重要程度和隐因子，提供的候选论文比其他算法更具有多样性。

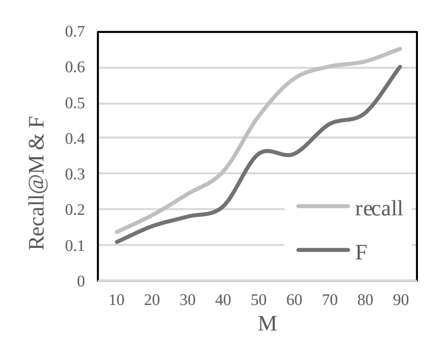


图 6　不同候选推荐数量对推荐结果的影响

对于用户历史数据数量，设置动态的候选推荐数量会对推荐结果产生影响。如图 5 所示，横坐标表示测试集中根据用户不同操作数量设置的候选推荐数量。实验过程设置混合模型的参数 $d=0.9$、$f=20$，候选推荐数量为用户操作的行为数据（超过 100 则设置为 100）。可以看出根据用户的操作数量选择候选推荐数量，不同数量的候选推荐对推荐性能的影响。用户的行为数据越多，越能找到用户的可能喜欢的论文。

4　结束语

本文提出一种分布式图计算的学术论文推荐算法。针对大数据时代提高个性化推荐方法的精度与效率的问题。首先用图模型来表示用户和论文间的关系，在论文关系中计算论文重要程度，并利用论文重要程度对用户论文关系进行加权处理。然后将基于图模型的 SVD++算法应用到用户-论文图模型中产生推荐。并且为了提高算法的可扩展的性，提出了图计算算法的分布式的实现。最后，实验结果验证了本文提出的混合算法模型有效性。

参考文献：

[1] DBLP. Statistics new records per year [DB/OL]. 2017. http://dblp. uni-trier. de/statistics/newrecordsperyear.

[2] Beel J, Langer S, Genzmehr M, *et al.* Research paper recommender system evaluation: a quantitative literature survey [C]// Proc of International Workshop on Reproducibility and Replication in Recommender Systems Evaluation. 2013: 15-22.

[3] Li G, Zhang Z, Wang L, *et al.* One-class collaborative filtering based on rating prediction and ranking prediction [J]. Knowledge-Based Systems, 2017, 124: 46-54.

[4] 高洁. 高质量学术资源推荐方法的研究与实现 [D]. 北京: 北京邮电大学, 2014.

[5] Amami M, Pasi G, Stella F, *et al.* An LDA-Based approach to scientific

paper recommendation [C]// Proc of International Conference on Applications of Natural Language to Information Systems. [S. l.] : Springer International Publishing, 2016: 200-210.

[6] Pan L, Dai X, Huang S, *et al.* Academic paper recommendation based on heterogeneous graph [J]. Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. 2015, 9427: 381-392.

[7] Habib R, Afzal M T. Paper recommendation using citation proximity in bibliographic coupling [J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2017, 25: 2708-2718.

[8] Amami M, Faiz R, Stella F, *et al.* A graph based approach to scientific paper recommendation [C]// Proc of International Conference. 2017: 777-782.

[9] Gonzalez J E, Low Y, Gu H, *et al.* PowerGraph: distributed graph-parallel computation on natural graphs [C]// Proc of Usenix Conference on Operating Systems Design and Implementation. [S. l] : USENIX Association, 2012: 17-30.

[10] Sugiyama K, Kan M Y. Towards higher relevance and serendipity in scholarly paper recommendation [J]. ACM SIGWEB Newsletter, 2015 (4): 1-16.

[11] 张玉连，袁伟. 隐语义模型下的科技论文推荐 [J]. 计算机应用与软件, 2015 (2): 37-40.

[12] 刘剑波，杨健. 基于 SVD+与行为分析的社交推荐 [J]. 计算机应用, 2013, 33 (1): 82-86.

[13] 谢玮，沈一，马永征. 基于图计算的论文审稿自动推荐系统 [J]. 计算机应用研究, 2016, 33 (3): 798-801.

[14] Sugiyama K, Kan M Y. A comprehensive evaluation of scholarly paper recommendation using potential citation papers [J]. International Journal on Digital Libraries, 2015, 16 (2): 91-109.

[15] Shams B, Haratizadeh S. Graph-based collaborative ranking [M]. [S. l.] ; Pergamon Press, Inc. 2017.

[16] Gonzalez J E, Xin R S, Dave A, *et al.* GraphX: graph processing in a distributed dataflow framework [C]// Proc of USENIX Conference on Operating Systems Design and Implementation. [S. l.] : USENIX Association, 2014: 599-613.

[17] 何海洋. 基于矩阵分解及其图模型的协同过滤推荐算法研究 [D]. 西安：西安电子科技大学, 2015.